

Competition, Alignment, and Equilibria in Digital Marketplaces

Meena Jagadeesan, Michael I. Jordan, Nika Haghtalab

University of California, Berkeley
 mjagadeesan@berkeley.edu, jordan@cs.berkeley.edu, nika@berkeley.edu

Abstract

Competition between traditional platforms is known to improve user utility by aligning the platform’s actions with user preferences. But to what extent is alignment exhibited in data-driven marketplaces? To study this question from a theoretical perspective, we introduce a duopoly market where platform actions are bandit algorithms and the two platforms compete for user participation. A salient feature of this market is that the quality of recommendations depends on both the bandit algorithm and the amount of data provided by interactions from users. This interdependency between the algorithm performance and the actions of users complicates the structure of market equilibria and their quality in terms of user utility. Our main finding is that competition in this market does not perfectly align market outcomes with user utility. Interestingly, market outcomes exhibit misalignment not only when the platforms have separate data repositories, but also when the platforms have a shared data repository. Nonetheless, the data sharing assumptions impact what mechanism drives misalignment and also affect the specific form of misalignment (e.g. the quality of the best-case and worst-case market outcomes). More broadly, our work illustrates that competition in digital marketplaces has subtle consequences for user utility that merit further investigation.¹

1 Introduction

Recommendation systems are the backbone of numerous digital platforms—from web search engines to video sharing websites to music streaming services. To produce high-quality recommendations, these platforms rely on data which is obtained through interactions with users. This fundamentally links the quality of a platform’s services to how well the platform can attract users.

What a platform must do to attract users depends on the amount of competition in the marketplace. If the marketplace has a single platform—such as Google prior to Bing or Pandora prior to Spotify—then the platform can accumulate users by providing any reasonably acceptable quality of service given the lack of alternatives. This gives the platform great flexibility in its choice of recommendation algorithm. In contrast, the presence of competing platforms makes user

participation harder to achieve and intuitively places greater constraints on the recommendation algorithms. This raises the questions: *how does competition impact the recommendation algorithms chosen by digital platforms? How does competition affect the quality of service for users?*

Conventional wisdom tells us that competition benefits users. In particular, users vote with their feet by choosing the platform on which they participate. The fact that users have this power forces the platforms to fully cater to user choices and thus improves user utility. This phenomenon has been formalized in classical markets where firms produce homogeneous products (Baye and Kovenock 2008), where competition has been established to perfectly align market outcomes with user utility. Since user wellbeing is central to the healthiness of a market, perfect competition is traditionally regarded as the “gold standard” for a healthy marketplace: this conceptual principle underlies measures of market power (Lerner 1934) and antitrust policy (Gellhorn 1975).

In contrast, competition has an ambiguous relationship with user wellbeing in digital marketplaces, where digital platforms are data-driven and compete via recommendation algorithms that rely on data from user interactions. Informally speaking, these marketplaces exhibit an interdependency between user utility, the platforms’ choices of recommendation algorithms, and the collective choices of other users. In particular, the size of a platform’s user base impacts how much data the platform has and thus the quality of its service; as a result, an individual user’s utility level depends on the number of users that the platform has attracted thus far. Having a large user base enables a platform to have an edge over competitors without fully catering to users, which casts doubt on whether classical alignment insights apply to digital marketplaces.

The ambiguous role of competition in digital marketplaces—which falls outside the scope of our classical understanding of competition power—has gained center stage in policymaking discourse. Indeed, several policy reports (Stigler Committee 2019; Crémer, de Montjoye, and Schweitzer 2019) have been dedicated to highlighting ways in which the structure of digital marketplaces differs from that of classical markets. For example, these reports suggest that data accumulation can encourage market tipping, leaving users vulnerable to harm (as we discuss at the end of Section 1.1). Yet, no theoretical foundation has emerged to formally examine the market structure of digital

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹See <https://arxiv.org/abs/2208.14423> for the full version.

marketplaces and assess potential interventions.

1.1 Our Contributions

Our work takes a step towards building a theoretical foundation for studying competition in digital marketplaces. We present a framework for studying platforms that compete on the basis of learning algorithms, focusing on alignment with user utility at equilibrium. We consider a stylized duopoly model based on a multi-armed bandit problem where user utility depends on the incurred rewards. We show that *competition may no longer perfectly align market outcomes with user utility*. Interestingly, there can be multiple equilibria, and the gap between the best equilibria and the worst equilibria can be substantial.

Model. We consider a market with two platforms and a population of users. Each platform selects a bandit algorithm from a class \mathcal{A} . After the platforms commit to algorithms, each user decides which platform to participate on. Each user’s utility is the (potentially discounted) cumulative reward that they receive from the bandit algorithm of the platform that they chose. Users arrive at a Nash equilibrium. Each platform’s utility is the number of users who participate on that platform, and the platforms arrive at a Nash equilibrium. The platforms either maintain *separate data repositories* about the rewards of their own users, or the platforms maintain a *shared data repository* about the rewards of all users.

Alignment results. To formally consider alignment, we introduce a metric—that we call the *user quality level*—that captures the utility that a user would receive when a given pair of competing bandit algorithms are implemented and user choices form an equilibrium. Table 1 summarizes the alignment results in the case of a single user and multiple users. A key quantity that appears in the alignment results is $R_{A'}(n)$, which denotes the expected utility that a user receives from the algorithm A' when n users all participate in the same algorithm.

For the case of a single user, an *idealized* form of alignment holds: the user quality level at any equilibrium is the optimal utility $\max_{A'} R_{A'}(1)$ that a user can achieve within the class of algorithms \mathcal{A} . Idealized alignment holds regardless of the informational assumptions on the platform.

The nature of alignment fundamentally changes when there are multiple users. At a high level, we show that idealized alignment breaks down since the user quality level is no longer guaranteed to be the global optimum, $\max_{A'} R_{A'}(N)$, that cooperative users can achieve. Nonetheless, a weaker form of alignment holds: the user quality level nonetheless never falls below the single-user optimum $\max_{A'} R_{A'}(1)$.

More formally, consider the setting where the platforms have separate data repositories. We show that there can be many qualitatively different Nash equilibria for the platforms. The user quality level across all equilibria actually spans the full set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$; i.e., any user quality level is realizable in some Nash equilibrium of the platforms and its associated Nash equilibrium of the users (Theorem 2). Moreover, the user quality level at any equilibrium is contained in the set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ (Theorem 3). When the number of users N is large, the gap between $\max_{A'} R_{A'}(1)$ and $\max_{A'} R_{A'}(N)$ can be signifi-

cant since the latter is given access to N times as much data at each time step than the former. The intuition behind this result is that the performance of an algorithm is controlled not only by its efficiency in transforming information to action, but also by the level of data it has gained through its user base. Since platforms have separate data repositories, a platform can thus make up for a suboptimal algorithm by gaining a significant user base.

What if the platforms were to share data? At first glance, it might appear that with data sharing, a platform can no longer make up for a suboptimal algorithm with data, and the idealized form of alignment would be recovered. However, we construct two-armed bandit problem instances where every symmetric equilibrium for the platforms has user quality level strictly below the global optimal $\max_{A'} R_{A'}(N)$ (Theorems 4-5). The mechanism for this suboptimality is that the global optimal solution requires “too much” exploration. If other users engage in their “fair share” of exploration, an individual user would prefer to explore less and free-ride off of the data obtained by other users. To formalize this, we establish a connection to strategic experimentation (Bolton and Harris 1999). We further show that although all of the user quality levels in $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ may not be realizable, the user quality level at any symmetric equilibria is guaranteed to be within this set (Theorem 6).

Connection to policy reports. Our work provides a mathematical explanation of phenomena documented in policy reports (Stigler Committee 2019; Crémer, de Montjoye, and Schweitzer 2019). The first phenomena that we consider is *market dominance from data accumulation*. The accumulation of data has been suggested to result in winner-takes-all markets where a single player dominates and where market entry is challenging (Stigler Committee 2019). The data advantage of the dominant platform can lead to lower quality services and lower user utility. Theorems 2-3 formalize this mechanism (see also the discussion in Section 4.2). The second phenomena that we consider is the *impact of shared data access*. While the separate data setting captures much of the status quo of proprietary data repositories, sharing data access has been proposed as a solution to market dominance (Crémer, de Montjoye, and Schweitzer 2019). Will shared data access deliver on its promises? Theorems 4-5 show sharing data does not solve alignment issues.

1.2 Related Work

We discuss the relation to research on *competing platforms*, *incentivizing exploration*, and *strategic experimentation*.

Competing platforms. Aridor et al. (2020) examine the interplay between competition and exploration in a duopoly bandit setup with fully myopic users. They show that platforms must both choose a greedy algorithm at equilibrium and thus illustrate that competition is at odds with regret minimization. In contrast, we take a user-centric perspective and investigate alignment with user utility. Interestingly, the result in Aridor et al. (2020) can be viewed as alignment (since the optimal choice for a fully myopic user results in regret), and our results similarly recover idealized alignment in the special case when users are fully myopic. Going beyond the setup of Aridor et al. (2020), we investigate non-myopic

	Single user	Multiple users
Separate data repositories	$\max_{A'} R_{A'}(1)$	$[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$
Shared data repository	$\max_{A'} R_{A'}(1)$	subset of $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ (<i>strict</i> subset in safe-risky arm problem)

Table 1: User quality level of the Nash equilibrium for the platforms. A marketplace with a single user exhibits *idealized alignment*, where the user quality level is maximized. A marketplace with multiple users can have equilibria with a vast range of user quality levels, and there are subtle differences between the separate and shared data settings.

users and allow multiple users to arrive at every round, and we show that alignment breaks down in this general setting.

Outside of the bandits framework, another line of work has also studied the behavior of competing learners when users can choose between platforms. Ben-Porat and Tennenholtz (2017, 2019) study equilibrium predictors chosen by competing *offline* learners in a PAC learning setup. Other work has focused on the dynamics when multiple learners apply out-of-box algorithms, showing that specialization can emerge (Ginart et al. 2021; Dean et al. 2022) and examining the role of data purchase (Kwon, Ginart, and Zou 2022); however, these works do not consider which algorithms the learners are incentivized to choose to gain users. In contrast, we investigate *equilibrium* algorithms chosen by *online* learners.

Other aspects of competing platforms that have been studied include competition under exogeneous network effects (Rysman 2009; Weyl and White 2014), experimentation in price competition (Bergemann and Välimäki 2000), dueling algorithms which compete for a single user (Immorlica et al. 2011), competition when firms select innovation levels whose cost decreases with more data (Prüfer and Schottmüller 2021), and measures of a digital platform’s power in a marketplace (Hardt, Jagadeesan, and Mendler-Dünner 2022).

Incentivizing exploration. This line of work has examined how the availability of outside options impacts bandit algorithms. Kremer, Mansour, and Perry (2013) show that Bayesian Incentive Compatibility (BIC) suffices to guarantee that users will stay on the platform. Follow-up work (e.g., (Mansour, Slivkins, and Syrgkanis 2015; Sellke and Slivkins 2021)) examines what bandit algorithms are BIC. Frazier et al. (2014) explore the use of monetary transfers.

Strategic experimentation. This line of work has investigated equilibria when a population of users each choose a bandit algorithm. Bolton and Harris (1999, 2000) analyze equilibria in a risky-safe arm bandit problem: we leverage their results in our analysis of the shared data setting. Strategic experimentation (see (Hörner and Skrzypacz 2017) for a survey) has investigated exponential bandit problems (Keller, Rady, and Cripps 2005), the impact of observing actions instead of payoffs (Rosenberg, Solan, and Vieille 2007), and the impact of cooperation (Brânzei and Peres 2021).

2 Model

We consider a duopoly market with two platforms performing a multi-armed bandit learning problem and a population of N users, u_1, \dots, u_N , who choose between platforms. Platforms

commit to bandit algorithms, and then each user chooses a single platform to participate on for the learning task.

2.1 Multi-Armed Bandit Setting

Consider a Bayesian bandit setting where there are k arms with priors $\mathcal{D}_1^{\text{Prior}}, \dots, \mathcal{D}_k^{\text{Prior}}$. At the beginning of the game, the mean rewards of arms are drawn from the priors $r_1 \sim \mathcal{D}_1^{\text{Prior}}, \dots, r_k \sim \mathcal{D}_k^{\text{Prior}}$. These mean rewards are unknown to both the users and the platforms but are shared across the two platforms. If the user’s chosen platform recommends arm i , the user receives reward drawn from a noisy distribution $\mathcal{D}^{\text{Noise}}(r_i)$ with mean r_i .

Let \mathcal{A} be a class of bandit algorithms that map the information state given by the posterior distributions to an arm to be pulled. The *information state* $\mathcal{I} = [\mathcal{D}_1^{\text{Post}}, \dots, \mathcal{D}_k^{\text{Post}}]$ is taken to be the set of posterior distributions for the mean rewards of each arm. Each algorithm $A \in \mathcal{A}$ is represented as a function mapping the information state \mathcal{I} to a distribution over $[k]$.

Running example: risky-safe arm bandit problem. To concretize our results, we consider the *risky-safe arm bandit problem* as a running example. The noise distribution $\mathcal{D}^{\text{Noise}}(r_i)$ is a Gaussian $N(r_i, \sigma^2)$. The first arm is a *risky arm* whose prior distribution $\mathcal{D}_1^{\text{Prior}}$ is over the set $\{l, h\}$, where l corresponds to a “low reward” and h corresponds to a “high reward.” The second arm is a *safe arm* with known reward $s \in (l, h)$ (the prior $\mathcal{D}_2^{\text{Prior}}$ is a point mass at s). In this case, the information state \mathcal{I} permits a one-dimensional representation given by $p(\mathcal{I}) := \mathbb{P}_{X \sim \mathcal{D}_1^{\text{Post}}}[X = h]$.

We construct a natural algorithm class as follows. For a measurable function $f : [0, 1] \rightarrow [0, 1]$, let A_f be the associated algorithm defined so $A_f(\mathcal{I})$ is a distribution that is 1 with probability $f(p(\mathcal{I}))$ and 2 with probability $1 - f(p(\mathcal{I}))$. We define $\mathcal{A}_{\text{all}} := \{A_f \mid f : [0, 1] \rightarrow [0, 1] \text{ is measurable}\}$ to be the class of all randomized algorithms. This class contains Thompson sampling ($A_{f_{\text{TS}}}$ is given by $f_{\text{TS}}(p) = p$), the Greedy algorithm ($A_{f_{\text{Greedy}}}$ is given by $f_{\text{Greedy}}(p) = 1$ if $ph + (1 - p)l \geq s$ and $f_{\text{Greedy}}(p) = 0$ otherwise), and mixtures of these algorithms with uniform exploration.

2.2 Platforms, Users, and Data

The interactions between the platform and users impact the data that the platform receives for its learning task. The platform action space \mathcal{A} is a class of bandit algorithms that map an information state \mathcal{I} to an arm to be pulled. The user action space is $\{1, 2\}$. For $1 \leq i \leq N$, we denote by $p^i \in \{1, 2\}$ the action chosen by user u_i .

Order of play. The platforms commit to algorithms A_1 and A_2 respectively, and then users simultaneously choose their actions p^1, \dots, p^N prior to the beginning of the learning task. We emphasize that user i participates on platform p^i for the *full duration* of the learning task. (In the full version, we discuss the assumption that users cannot switch platforms between time steps.)

Data sharing assumptions. In the *separate data repositories* setting, each platform has its own (proprietary) data repository for keeping track of the rewards incurred by its own users. Platforms 1 and 2 thus have separate information states given by $\mathcal{I}_1 = [\mathcal{D}_{1,1}^{\text{Post}}, \dots, \mathcal{D}_{1,k}^{\text{Post}}]$ and $\mathcal{I}_2 = [\mathcal{D}_{2,1}^{\text{Post}}, \dots, \mathcal{D}_{2,k}^{\text{Post}}]$, respectively. In the *shared data repository* setting, the platforms share an information state $\mathcal{I}_{\text{shared}} = [\mathcal{D}_1^{\text{Post}}, \dots, \mathcal{D}_k^{\text{Post}}]$, which is updated based on the rewards incurred by users of both platforms.²

Learning task. The learning task is determined by the choice of A_1 and A_2 , user actions p^1, \dots, p^N , and specifics of data sharing between platforms. At each time step:

1. Each user u_i arrives at platform p^i . The platform p^i recommends arm $a_i \sim A_i(\mathcal{I})$ to that user, where \mathcal{I} denotes the information state of the platform. (The randomness of arm selection is fully independent across users and time steps.) The user u_i receives noisy reward $\mathcal{D}^{\text{Noise}}(r_{a_i})$.
2. After providing recommendations to all of its users, platform 1 observes the rewards incurred by users in $S_1 := \{i \in [N] \mid p^i = 1\}$. Platform 2 similarly observes the rewards incurred by users in $S_2 := \{i \in [N] \mid p^i = 2\}$. Each platform then updates their information state \mathcal{I} with the corresponding posterior updates.
3. A platform may have access to external data that does not come from users. To capture this, we introduce background information into the model. Both platforms observe the same *background information* of quality $\sigma_b \in (0, \infty]$. For each arm i , the platforms observe the same realization of a noisy reward $\mathcal{D}^{\text{Noise}}(r_i)$.

In other words, platforms receive information from users (and background information), and users receive rewards based on the recommendations of the platform that they have chosen.

2.3 Utility Functions and Equilibrium Concept

User utility is generated by rewards, while the platform utility is generated by *user participation*.

User utility function. We follow the standard discounted formulation for bandit problems (e.g. (Gittins and Jones 1979; Bolton and Harris 1999)), where the utility incurred by a user is defined by the expected (discounted) cumulative reward received across time steps. The discount factor β parameterizes the extent to which agents are myopic. Let $\mathcal{U}(p^i; \mathbf{p}^{-i}, A_1, A_2)$ denote the utility of a user u_i if they take action $p^i \in \{1, 2\}$ when other users take actions $\mathbf{p}^{-i} \in \{1, 2\}^{N-1}$ and the platforms choose A_1 and A_2 . For clarity, we make this explicit in the case of discrete time setup

²In web search, recommender systems can query each other, effectively building a shared information state.

with horizon length $T \in [1, \infty]$. Let $a_i^t = a_i^t(A_1, A_2, \mathbf{p})$ denote the arm recommended to user u_i at time step t . The utility is defined to be $\mathcal{U}(p^i; \mathbf{p}^{-i}, A_1, A_2) := \mathbb{E} \left[\sum_{t=1}^T \beta^t r_{a_i^t} \right]$, where the expectation is over randomness of the incurred rewards and the algorithms. In the case of continuous time, the utility is $\mathcal{U}(p^i; \mathbf{p}^{-i}, A_1, A_2) := \mathbb{E} \left[\int e^{-\beta t} d\pi(t) \right]$, where the $\beta \in [0, \infty)$ denotes the discount factor and $d\pi(t)$ denotes the payoff received by the user. In both cases, observe that the utility function is *symmetric* in user actions.

The utility function implicitly differs in the separate and shared data settings, since the information state evolves differently in these two settings. When we wish to make this distinction explicit, we denote the corresponding utility functions by $\mathcal{U}^{\text{separate}}$ and $\mathcal{U}^{\text{shared}}$.

User equilibrium concept. After the platforms commit to algorithms A_1 and A_2 , the users end up at a pure strategy Nash equilibrium of the resulting game. More formally, let $\mathbf{p} \in \{1, 2\}^N$ be a *pure strategy Nash equilibrium for the users* if $p_i \in \arg \max_{p \in \{0,1\}} \mathcal{U}(p; \mathbf{p}^{-i}, A_1, A_2)$ for all $1 \leq i \leq N$. The existence of a pure strategy Nash equilibrium follows from the assumption that the game is symmetric and the action space has 2 elements (Cheng et al. 2004).

One subtlety is that there can be multiple equilibria in this general-sum game. For example, there are at least 2 (pure strategy) equilibria when platforms commit to the same algorithm—one equilibrium where all users choose the first platform, and another where all users choose the second platform. We denote by \mathcal{E}_{A_1, A_2} the set of pure strategy Nash equilibria when the platforms choose algorithms A_1 and A_2 . We simplify the notation and use \mathcal{E} when A_1 and A_2 are clear from the context. In the full version of the paper, we discuss our choice of solution concept in greater detail.

Platform utility and equilibrium concept. The utility of the platform roughly corresponds to the number of users who participate on that platform. This captures that in markets for digital goods, where platform revenue is often derived from advertisement or subscription fees, the number of users serviced is a proxy for platform revenue.

When formalizing platform utility, the fact that there can be several user equilibria for a given choice of platform algorithms creates ambiguity. To resolve this, we consider the worst-case user equilibrium for the platform, and we define platform utility to be the *minimum number of users that a platform would receive at any pure strategy equilibrium for the users*. More formally, when platform 1 chooses A_1 and platform 2 chooses A_2 , the utilities are given by:

$$v_1(A_1; A_2) := \min_{\mathbf{p} \in \mathcal{E}} \sum_{i=1}^N 1[p^i = 1], \quad v_2(A_2; A_1) = \min_{\mathbf{p} \in \mathcal{E}} \sum_{i=1}^N 1[p^i = 2]. \quad (1)$$

The equilibrium concept for the platforms is a *pure strategy Nash equilibrium*, and we often focus on *symmetric* equilibria. We discuss the existence of such an equilibrium in Sections 4–5. We note that at equilibrium, the utility for the platforms is typically 0. Platforms earning zero equilibrium utility in our model mirrors firms earning zero equilibrium profit in price competition (Baye and Kovenock 2008). However, there is an important distinction: platform utility *ex-post* (after users

choose between platforms) may no longer be 0 and may be as large as N , while firm profit in price competition is 0 ex-post.

3 The Alignment of a Market Outcome

The *alignment* of an equilibrium outcome for the platforms is measured by the amount of user utility that it generates. In Section 3.1 we introduce the *user quality level* to formalize alignment. In Section 3.2, we show an idealized form of alignment for $N = 1$ (Theorem 1). In Section 3.3, we turn to the case of multiple users and discuss benchmarks for the user quality level. In Section 3.4, we describe mild assumptions on \mathcal{A} that we use in our alignment results for multiple users.

3.1 User Quality Level

Given platform algorithms $A_1 \in \mathcal{A}$ and $A_2 \in \mathcal{A}$, we introduce the *user quality level* $Q(A_1, A_2)$ to capture the utility that a user would receive when the platforms choose A_1 and A_2 and when user choices form an equilibrium. When formalizing this, the potential multiplicity of user equilibria creates ambiguity (like in the definition of platform utility in (1)), and different users potentially receiving different utilities creates further ambiguity. We again take a worst-case perspective and formalize the user quality level as the minimum over equilibria $\mathbf{p} \in \mathcal{E}$ and over users $1 \leq i \leq N$.

Definition 1 (User quality level). *Given algorithms A_1 and A_2 chosen by the platforms, the user quality level is defined to be $Q(A_1, A_2) := \min_{\mathbf{p} \in \mathcal{E}, 1 \leq i \leq N} \mathcal{U}(p_i; \mathbf{p}^{-i}, A_1, A_2)$.*

As we discuss in the full version, our insights about alignment would remain unchanged if we were to define the user quality level based on an arbitrary user equilibrium and user, rather than taking a minimum.

To simplify notation in our alignment results, we introduce the *reward function* which captures how the utility that a given algorithm generates changes with the number of users who contribute to its data repository. For an algorithm $A \in \mathcal{A}$, let the reward function $R_A : [N] \rightarrow \mathbb{R}$ be defined by:

$$R_A(n) := \mathcal{U}^{\text{separate}}(1; \mathbf{p}_{n-1}, A, A),$$

where \mathbf{p}_{n-1} is a vector with $n - 1$ coordinates equal to one.

3.2 Idealized Alignment Result: A Single User

When there is a single user, the platform algorithms turn out to be perfectly aligned with user utilities at equilibrium. To formalize this, we consider the optimal utility that could be obtained by a user across any choice of actions by the platforms and users (not necessarily at equilibrium): that is, $\max_{p \in \{1, 2\}, A_1 \in \mathcal{A}, A_2 \in \mathcal{A}} \mathcal{U}(p; \emptyset, A_1, A_2)$. Using the setup of the single-user game, we can see that this is equal to $\max_{A \in \mathcal{A}} \mathcal{U}(1; \emptyset, A, A) = \max_{A \in \mathcal{A}} R_A(1)$. We show that the user quality level always meets this benchmark.

Theorem 1. *Suppose that $N = 1$, and consider either the separate data setting or the shared data setting. If (A_1, A_2) is a pure strategy Nash equilibrium for the platforms, then the user quality level $Q(A_1, A_2)$ is equal to $\max_{A \in \mathcal{A}} R_A(1)$.*

Theorem 1 shows that in a single-user market, two firms is sufficient to perfectly align firm actions with user utility—this stands in parallel to classical Bertrand competition in the pricing setting (Baye and Kovenock 2008).

3.3 Benchmarks for User Quality Level

In the case of multiple users, this idealized form of alignment turns out to break down, and formalizing alignment requires a nuanced consideration of benchmarks. We define the *single-user optimal utility* to be $\max_{A \in \mathcal{A}} R_A(1)$. This corresponds to maximal possible user utility that can be generated by a platform who only thus relies on a single user for all of its data. On the other hand, we define the *global optimal utility* to be $\max_{A \in \mathcal{A}} R_A(N)$. This corresponds to the maximal possible user utility that can be generated by a platform when all of the users are forced to be on that platform.

3.4 Assumptions on \mathcal{A}

When there are multiple users, we require mild assumptions on \mathcal{A} . *Information monotonicity* requires that an algorithm A 's performance in terms of user utility does not worsen with additional posterior updates to the information state. Our first two instantiations—strict information monotonicity and information constantness—require that the user utility of A grow monotonically in the number of other users participating in the algorithm. Our third instantiation—side information monotonicity—requires that the user utility of A not decrease if other users also update the information state.

Assumption 1 (Information monotonicity). *For any given discount factor β and number of users N , an algorithm $A \in \mathcal{A}$ is **strictly information monotonic** if $R_A(n)$ is strictly increasing in n for $1 \leq n \leq N$. An algorithm A is **information constant** if $R_A(n)$ is constant in n for $1 \leq n \leq N$. An algorithm A is **side information monotonic** if for every measurable function f mapping information states to distributions over $[k]$ and for every $1 \leq n \leq N - 1$, it holds that $\mathcal{U}^{\text{shared}}(1; \mathbf{2}_n, A, f) \geq R_A(1)$ where $\mathbf{2}_n \in \{1, 2\}^n$ has all coordinates equal to 2.*

Utility richness requires that the set of user utilities spanned by \mathcal{A} is a sufficiently rich interval.

Assumption 2 (Utility richness). *A class of algorithms \mathcal{A} is **utility rich** if the set of utilities $\{R_A(N) \mid A \in \mathcal{A}\}$ is a contiguous set, the supremum of $\{R_A(N) \mid A \in \mathcal{A}\}$ is achieved, and there exists $A' \in \mathcal{A}$ such that $R_{A'}(N) \leq \max_{A \in \mathcal{A}} R_A(1)$.*

Discussion of assumptions. Utility richness holds (almost) without loss of generality, by taking the closure of an algorithmic class under mixtures with exploration. However, not all algorithms are information monotone. We nevertheless show the information monotonicity of several algorithms for the risky-safe arm setup, including any algorithm under undiscounted rewards and Thompson sampling under discounted rewards. These results are of independent interest and are stated in the full version. More broadly, understanding information monotonicity is crucial for studying the incentive properties of bandit algorithms: indeed prior work (e.g. Aridor et al. (2020)) has explored variants of this assumption.

4 Separate Data Repositories

We investigate alignment when the platforms have separate data repositories. In Section 4.1, we show that there can be many qualitatively different equilibria for the platforms

and characterize their alignment. In Section 4.2, we discuss factors that drive the level of misalignment in a marketplace.

4.1 Multitude of Equilibria and Alignment

In contrast with the single user setting, the marketplace can exhibit multiple equilibria for the platforms. As a result, to investigate alignment, we investigate the *range* of achievable user quality levels. Our main finding is that the equilibria can exhibit a vast range of alignment properties. In particular, *every* user quality level in between the single-user optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(1)$ and the global optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(N)$ can be realized by some equilibrium.

Theorem 2. *Suppose that each algorithm in \mathcal{A} is either strictly information monotonic or information constant (Assumption 1), and suppose that \mathcal{A} is utility rich (Assumption 2). For every $\alpha \in [\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N)]$, there exists a symmetric pure strategy Nash equilibrium (A, A) in the separate data setting such that $Q(A, A) = \alpha$.*

Nonetheless, there is a baseline (although somewhat weak) form of alignment achieved by all equilibria. In particular, every equilibrium for the platforms has user quality level at least the single-user optimum $\max_{A' \in \mathcal{A}} R_{A'}(1)$.

Theorem 3. *Suppose that each algorithm in \mathcal{A} is either strictly information monotonic or information constant (see Assumption 1). In the separate data setting, at any pure strategy Nash equilibrium (A_1, A_2) for the platforms, the user quality level lies in the following interval:*

$$Q(A_1, A_2) \in \left[\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N) \right].$$

An intuition for these results is that the performance of an algorithm depends not only on how it transforms information to actions, but also on the amount of information to which it has access. A platform can make up for a suboptimal algorithm by attracting a significant user base: if a platform starts with the full user base, *no single user* may want to switch to the competing platform, even if the competing platform chooses a stricter better algorithm. However, if a platform’s algorithm is highly suboptimal, then the competing platform will indeed be able to win the full user base.

Proof sketches of Theorem 2 and Theorem 3. We first show pure strategy equilibria for users take a simple form. In particular, we show that every pure strategy equilibrium $p^* \in \mathcal{E}^{A_1, A_2}$ is in the set $\{[1, \dots, 1], [2, \dots, 2]\}$. The reward functions $R_{A_1}(\cdot)$ and $R_{A_2}(\cdot)$ determine which of these two solutions are in \mathcal{E}^{A_1, A_2} .

This characterization of the set \mathcal{E}^{A_1, A_2} enables us to reason about the platform equilibria. To prove Theorem 2, the key ingredient is that (A, A) is an equilibrium for the platforms as long as $R_A(N) \geq \max_{A'} R_{A'}(1)$. To prove Theorem 3, the key insight is that platforms can’t both choose highly suboptimal algorithms at equilibrium.

4.2 What Drives the Level of Misalignment?

The multiplicity of equilibria makes it subtle to reason about the alignment in a marketplace. The level of misalignment depends on two factors: first, the size of the range of realizable

user quality levels, and second, the selection of equilibrium within this range. We explore each of these factors.

How large is the range of possible user quality levels?

Both the algorithm class and the structure of the user utility function determine the size of the range of possible user quality levels. We informally examine the role of the user’s discount factor on the size of this range.

First, consider the case where users are fully non-myopic. The gap between the single-user optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(1)$ and global optimal utility $\max_{A' \in \mathcal{A}} R_{A'}(N)$ can be substantial. For example, consider an algorithm A' whose regret grows according to \sqrt{T} where T is the number of samples collected, and let $\text{OPT} := \mathbb{E}_{r_1 \sim \mathcal{D}_1^{\text{prior}}, \dots, r_k \sim \mathcal{D}_k^{\text{prior}}} [\max_{1 \leq i \leq k} r_i]$ be the expected maximum reward of any arm. Since utility and regret are related up to additive factors for fully non-myopic users, we know $R_{A'}(1) \approx \text{OPT} - \sqrt{T}$ while $R_{A'}(N) \approx \text{OPT} - \sqrt{NT}$.

At the other extreme, consider all fully myopic users. The range collapses to a *single point*: in particular, $R_{A'}(1)$ is equal to $R_{A'}(N)$ for any algorithm $A' \in \mathcal{A}$. To see this, we observe that user utility is fully determined by the algorithm’s behavior at the first time step, which is determined before it receives any information from users. When users are partially non-myopic, the range is no longer a single point, but is intuitively smaller than in the undiscounted case.

Which equilibrium arises in a marketplace? When the gap between the single-user optimal and global optimal utility levels is substantial, it becomes ambiguous what user quality level will be realized in a given marketplace. Which equilibria arises in a marketplace depends on several factors.

One factor is the secondary aspects of the platform objective that aren’t fully captured by the number of users. For example, suppose that the platform derives other sources of revenue from recommending certain types of content (e.g. from recommending advertisements). If these additional sources of revenue are not aligned with user utility, then this could drive the marketplace towards lower user quality levels.

Another factor is the mechanism under which platforms arrive at equilibrium solutions. We informally show that market entry can result in the *worst possible user utility*. When one platform enters the marketplace before another platform, all users will initially choose the first platform. The first platform can lose users only if $R_{A_1}(N)$ is below the single-user optimal. Thus, the worst possible user quality level can arise, and this problem only worsens if the first platform accumulates data beforehand. This finding illustrates barriers to entry in digital marketplaces (Stigler Committee 2019).

5 Shared Data Repository

What happens when data is shared between the platforms? We show that both the nature of alignment and the forces that drive misalignment change. In Section 5.1, we show a construction where the user quality levels do not span the full set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$. Despite this, in Section 5.2, we establish that the user quality level at any symmetric equilibrium continues to be at least $\max_{A'} R_{A'}(1)$.

5.1 Construction Where Optimal Is Not Realizable

In contrast with the separate data setting, the set of user quality levels at symmetric equilibria for the platforms does not necessarily span the full set $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$. To demonstrate this, we construct bandit setups where every symmetric equilibrium (A, A) has user quality level $Q(A, A)$ below $\max_{A'} R_{A'}(N)$.

Theorem 4. *Let the algorithm class $\mathcal{A}_{all}^{cont} \subseteq \mathcal{A}_{all}$ consist of the algorithms A_f where $f(0) = 0$, $f(1) = 1$, and f is continuous at 0 and 1. In the shared data setting, there exists an undiscounted risky-safe arm bandit setup such that the set of realizable user quality levels for algorithm class \mathcal{A}_{all}^{cont} is equal to a singleton set:*

$$\{Q(A, A) \mid (A, A) \text{ an equilibrium for the platforms}\} = \{\alpha^*\}$$

where $\max_{A' \in \mathcal{A}} R_{A'}(1) < \alpha^* < \max_{A' \in \mathcal{A}} R_{A'}(N)$.

Theorem 5. *In the shared data setting, for any discount factor $\beta \in (0, \infty)$, there exists a risky-safe arm bandit setup such that the set of realizable user quality levels for algorithm class \mathcal{A}_{all} is equal to a singleton set:*

$$\{Q(A, A) \mid (A, A) \text{ an equilibrium for the platforms}\} = \{\alpha^*\}$$

where $\max_{A' \in \mathcal{A}} R_{A'}(1) \leq \alpha^* < \max_{A' \in \mathcal{A}} R_{A'}(N)$.

Theorems 4 and 5 illustrate examples where there is no symmetric equilibrium for the platforms that realizes the global optimal utility $\max_{A'} R_{A'}(N)$ —regardless of whether users are fully non-myopic or discounted. These results have interesting implications for shared data access as an intervention in digital marketplace regulation (e.g. see (Cr mer, de Montjoye, and Schweitzer 2019)). At first glance, it would appear that data sharing would resolve the alignment issues, since it prevents market dominance from data accumulation. However, our results illustrate that the platforms may still not align their actions with user utility at equilibrium.

Comparison of separate and shared data settings. We show fundamental differences in alignment when the platforms have a shared data repository and have separate data repositories. We focus on the undiscounted setup analyzed in Theorem 4; in this case, the algorithm class \mathcal{A}_{all}^{cont} satisfies information monotonicity and utility richness so the results in Section 4.1 are also applicable.

The first difference is that there is a unique symmetric equilibrium for the shared data setting, which stands in contrast to the range of equilibria for the separate data setting. Thus, while the particularities of equilibrium selection impact alignment in the separate data setting (see Section 4.2), these particularities are irrelevant for the shared data setting.

The second difference is that the user quality level in the shared data setting is strictly within the range of realizable user quality levels for the separate data setting. Alignment in the shared data setting is *strictly better* than the alignment of the worst possible equilibrium in the separate data setting. On the other hand, alignment in the shared data setting is *strictly worse* than the alignment of the best possible equilibrium in the separate data setting. This means that data sharing is worse for users than maintaining separate data and applying a well-designed equilibrium selection mechanism.

Mechanism for misalignment. Perhaps counterintuitively, the mechanism for misalignment in the shared data setting is that a platform must perfectly align its choice of algorithm with the preferences of a user (given the choices of other users). In particular, the algorithm that is optimal for one user given the actions of other users is different from the algorithm that would be optimal if the users were to cooperate. This is because exploration is costly to users, so users don't want to perform their fair share of exploration, and would rather free-ride off of the exploration of other users. A platform who chooses an algorithm with the global optimal strategy thus cannot maintain its user base. We formalize this phenomena by establishing a connection with *strategic experimentation*, drawing upon (Bolton and Harris 1999, 2000).

Proof sketches of Theorem 4 and Theorem 5. We relate the equilibria of our game to the equilibria of the following game G . Let G be an N user game where each user chooses an algorithm in \mathcal{A} within the same bandit problem setup as in our game. The users share an information state \mathcal{I} . At each time step, all N users arrive at the platform. Each user i pulls the arm drawn from $A_i(\mathcal{I})$ and updates \mathcal{I} . We show that the solution (A, A) is in equilibrium if and only if A is a symmetric pure strategy equilibrium of the game G .

To show Theorem 5, it suffices to analyze structure of the equilibria of G . The global optimal algorithm $A^* = \arg \max_{A' \in \mathcal{A}} R_{A'}(N)$ corresponds to the cooperative solution for the users, which may not be an equilibrium solution in G due “free-riding”. Interestingly, Bolton and Harris (1999, 2000)—in the context of strategic experimentation—formalized freeriding in a game similar to G . We provide a recap of the relevant aspects in the full version of the paper. We can adopt these results to analyze the equilibrium user utility in G . The full proof is deferred to the full version.

5.2 Weak Alignment

Although not all values in $[\max_{A'} R_{A'}(1), \max_{A'} R_{A'}(N)]$ can be realized, we show that the user quality level at any symmetric equilibrium is always at least $\max_{A'} R_{A'}(1)$.

Theorem 6. *Suppose that every algorithm in \mathcal{A} is side information monotonic (Assumption 1). In the shared data setting, at any symmetric equilibrium (A, A) , it holds that $Q(A, A) \in [\max_{A' \in \mathcal{A}} R_{A'}(1), \max_{A' \in \mathcal{A}} R_{A'}(N)]$.*

Theorem 6 demonstrates that free-riding cannot drive the user quality level below the single-user optimal. Theorem 6 parallels Theorem 3 from the separate data setting: in both cases, the market outcomes exhibit a weak form of alignment.

6 Discussion

We present a framework for analyzing competition between 2 multi-armed bandit learners interacting with a population of users. We analyze the *user quality level* to measure the alignment of market equilibria. We show that competition does not lead to perfect alignment, both when the platforms maintain separate data repositories and maintain a shared data repository. More broadly, our work provides a mathematical explanation of phenomena documented in policy reports and reveals that competition in data-driven marketplaces has subtle consequences that merit further inquiry.

Acknowledgments

We would like to thank Yannai Gonczarowski, Erik Jones, Rad Niazadeh, Jacob Steinhardt, Nilesh Tripuraneni, Abhishek Shetty, and Alex Wei for helpful comments on the paper. This work is in part supported by National Science Foundation under grant CCF-2145898, the Mathematical Data Science program of the Office of Naval Research under grant number N00014-18-1-2764, the Vannevar Bush Faculty Fellowship program under grant number N00014-21-1-2941, a C3.AI Digital Transformation Institute grant, the Paul and Daisy Soros Fellowship, and the Open Phil AI Fellowship.

References

- Aridor, G.; Mansour, Y.; Slivkins, A.; and Wu, Z. S. 2020. Competing Bandits: The Perils of Exploration Under Competition. *CoRR*, abs/2007.10144.
- Baye, M. R.; and Kovenock, D. 2008. *Bertrand Competition*. London: Palgrave Macmillan UK.
- Ben-Porat, O.; and Tennenholtz, M. 2017. Best Response Regression. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems (NIPS)*, 1499–1508.
- Ben-Porat, O.; and Tennenholtz, M. 2019. Regression Equilibrium. In *Proceedings of the 2019 ACM Conference on Economics and Computation (EC)*, 173–191.
- Bergemann, D.; and Välimäki, J. 2000. Experimentation in Markets. *The Review of Economic Studies*, 67(2): 213–234.
- Bolton, P.; and Harris, C. 1999. Strategic Experimentation. *Econometrica*, 67(2): 349–374.
- Bolton, P.; and Harris, C. 2000. Strategic Experimentation: the Undiscounted Case. *Incentives, Organizations and Public Economics – Papers in Honour of Sir James Mirrlees*, 53–68.
- Brânzei, S.; and Peres, Y. 2021. Multiplayer Bandit Learning, from Competition to Cooperation. In Belkin, M.; and Kpotufe, S., eds., *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, 679–723.
- Cheng, S.; Reeves, D. M.; Vorobeychik, Y.; and Wellman, M. P. 2004. Notes on equilibria in symmetric games. In *Proceedings of the 6th International Workshop on Game Theoretic and Decision Theoretic Agents (GTDT)*, 71–78.
- Crémer, J.; de Montjoye, Y.-A.; and Schweitzer, H. 2019. *Competition Policy for the digital era : Final report*. Publications Office of the European Union.
- Dean, S.; Curmei, M.; Ratliff, L. J.; Morgenstern, J.; and Fazel, M. 2022. Multi-learner risk reduction under endogenous participation dynamics. *CoRR*, abs/2206.02667.
- Frazier, P. I.; Kempe, D.; Kleinberg, J. M.; and Kleinberg, R. 2014. Incentivizing exploration. In Babaiouff, M.; Conitzer, V.; and Easley, D. A., eds., *ACM Conference on Economics and Computation, EC '14*, 5–22.
- Gellhorn, E. 1975. An Introduction to Antitrust Economics. *Duke Law Journal*, 1975(1): 1–43.
- Ginart, T.; Zhang, E.; Kwon, Y.; and Zou, J. 2021. Competing AI: How does competition feedback affect machine learning? In Banerjee, A.; and Fukumizu, K., eds., *The 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, 1693–1701.
- Gittins, J. C.; and Jones, D. M. 1979. A Dynamic Allocation Index for the Discounted Multiarmed Bandit Problem. *Biometrika*, 66(3): 561–565.
- Hardt, M.; Jagadeesan, M.; and Mendler-Dünner, C. 2022. Performative Power. *CoRR*, abs/2203.17232.
- Hörner, J.; and Skrzypacz, A. 2017. *Learning, Experimentation, and Information Design*, volume 1 of *Econometric Society Monographs*, 63–98. Cambridge University Press.
- Immorlica, N.; Kalai, A. T.; Lucier, B.; Moitra, A.; Postlewaite, A.; and Tennenholtz, M. 2011. Dueling algorithms. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, 215–224.
- Keller, G.; Rady, S.; and Cripps, M. 2005. Strategic Experimentation with Exponential Bandits. *Econometrica*, 73(1): 39–68.
- Kremer, I.; Mansour, Y.; and Perry, M. 2013. Implementing the "Wisdom of the Crowd". In Kearns, M. J.; McAfee, R. P.; and Tardos, É., eds., *Proceedings of the fourteenth ACM Conference on Electronic Commerce, EC 2013*, 605–606.
- Kwon, Y.; Ginart, A.; and Zou, J. 2022. Competition over data: how does data purchase affect users? *CoRR*, abs/2201.10774.
- Lerner, A. P. 1934. The Concept of Monopoly and the Measurement of Monopoly Power. *The Review of Economic Studies*, 1(3): 157–175.
- Mansour, Y.; Slivkins, A.; and Syrgkanis, V. 2015. Bayesian Incentive-Compatible Bandit Exploration. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC '15*, 565–582.
- Prüfer, J.; and Schottmüller, C. 2021. Competing with Big Data. *The Journal of Industrial Economics*, 69(4): 967–1008.
- Rosenberg, D.; Solan, E.; and Vieille, N. 2007. Social Learning in One-Arm Bandit Problems. *Econometrica*, 75(6): 1591–1611.
- Rysman, P. 2009. The Economics of Two-Sided Markets. *Journal of Economic Perspectives*, 23: 125–144.
- Sellke, M.; and Slivkins, A. 2021. The Price of Incentivizing Exploration: A Characterization via Thompson Sampling and Sample Complexity. In *EC '21: The 22nd ACM Conference on Economics and Computation*, 795–796.
- Stigler Committee. 2019. Final Report: Stigler Committee on Digital Platforms. <https://www.chicagobooth.edu/media/research/stigler/pdfs/digital-platforms---committee-report---stigler-center.pdf>. Accessed: 2023-4-26.
- Weyl, G.; and White, A. 2014. Let the Right 'One' Win: Policy Lessons from the New Economics of Platforms. *Competition Policy International*, 12: 29–51.